Orphanet Journal of
Rare Diseases

**RESEARCH**                                                                          **Open Access**

# DDIEM: drug database for inborn errors of metabolism

Marwa Abdelhakim[1,2], Eunice McMurray[3], Ali Raza Syed[1,2], Senay Kafkas[1,2], Allan Anthony Kamau[1], Paul N Schofield[3] and Robert Hoehndorf[1,2*]

## Abstract

**Background:** Inborn errors of metabolism (IEM) represent a subclass of rare inherited diseases caused by a wide range of defects in metabolic enzymes or their regulation. Of over a thousand characterized IEMs, only about half are understood at the molecular level, and overall the development of treatment and management strategies has proved challenging. An overview of the changing landscape of therapeutic approaches is helpful in assessing strategic patterns in the approach to therapy, but the information is scattered throughout the literature and public data resources.

**Results:** We gathered data on therapeutic strategies for 300 diseases into the Drug Database for Inborn Errors of Metabolism (DDIEM). Therapeutic approaches, including both successful and ineffective treatments, were manually classified by their mechanisms of action using a new ontology.

**Conclusions:** We present a manually curated, ontologically formalized knowledgebase of drugs, therapeutic procedures, and mitigated phenotypes. DDIEM is freely available through a web interface and for download at http://ddiem.phenomebrowser.net.

**Keywords:** Inborn errors of metabolism, Orphan disease, Ontology, Drug effect, Drug repositioning, Therapeutic procedure

## Background

Rare hereditary diseases together make up a significant component of overall morbidity and mortality around the world. A disease is considered rare in the United States (US) if it affects fewer than 200,000 individuals. With more than 7,000 rare diseases characterized, the overall population burden is therefore of the order of 20 to 30 million [1]. Consequently, it has become a significant concern for strategic health funding bodies in North America, Europe, and elsewhere to coordinate international infrastructure, research and development, and delivery of new therapeutics [2]. While much effort has gone into mobilizing patient data across national boundaries and undertaking extensive genetic studies, the Orphan drug initiatives in the US starting in 1983, together with ongoing programs at the FDA and NIH, as well as those funded by the European Commission and Canadian authorities, saw a dramatic increase in therapies for rare diseases. These were notable for a large number of drug repurposing successes, providing accelerated access to patients [3–5]. Currently, around 15% of new orphan drug approvals are for metabolic and endocrine therapies, most of these through the application of small molecules [6].

Metabolic hereditary diseases, or inborn errors of metabolism (IEMs), are a group of disorders that disrupt normal metabolism and physiology, ultimately affecting

*Correspondence: robert.hoehndorf@kaust.edu.sa
[1]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, 4700 KAUST, 23955 Thuwal, Kingdom of Saudi Arabia
[2]Computer, Electrical and Mathematical Sciences & Engineering Division (CEMSE), King Abdullah University of Science and Technology, 4700 KAUST,PO 23955 Thuwal, Saudi Arabia
Full list of author information is available at the end of the article

Abdelhakim *et al. Orphanet Journal of Rare Diseases*        (2020) 15:146

Page 2 of 12

almost all biochemical pathways and processes in the body. Individually metabolic diseases are rare, but, as with rare disease as a whole, relatively common when considered as a class of disease [7]. Over 1,000 distinct hereditary metabolic diseases have been identified and recently assigned to 130 categories in a new nosology [8], although the underlying causes are only understood for about half of these diseases.

Data resources have been developed for hereditary metabolic diseases. For example, the Rare Metabolic Diseases (RAMEDIS) database [9] contains information on rare diseases and their treatment options. However, RAMEDIS covers only 93 different IEMs and primarily relies on information from case reports and their associated available data on treatment options for most of the 93 disorders covered. IEMbase is a database that contains clinical, biochemical, and genetic information on 1,310 IEMs, as well as a nosology that classifies each disease into one of ten significant categories [10]. However, neither RAMEDIS nor IEMbase cover different treatment options and strategies for rare metabolic diseases.

Therapeutic approaches to inherited metabolic diseases are diverse and have to distinguish, for example, between loss of function, dominant gain of function, and the generation of toxic metabolic intermediates. Among therapeutic strategies that address metabolic diseases are gene therapy, metabolite level manipulation, transplantation or surgery, and small molecule therapies to stabilize or enhance residual enzyme activity. Therapies may address the underlying mechanism directly or indirectly. They may provide symptomatic or prophylactic therapy where some specific phenotypes of the syndrome are addressed, and sometimes they improve the long term pathological outcomes [11–13]. More importantly, a large number of rare IEMs do not currently have an approved treatment approach, and new therapeutic strategies are tested in individual cases or clinical trials, and are subsequently reported in the scientific literature [14]. An overview of the strategic landscape of therapeutic approaches to treating IEMs is of potential use not only for investigators developing new therapies but also for regulators and research funders.

We have developed the Drug Database for Inborn Errors of Metabolism (DDIEM), a database that covers experimental approaches, clinical treatments (established as well as investigational therapies) and their outcome, for diagnosed rare metabolic diseases. Individual diseases are linked to established information resources and databases. In order to classify therapeutic approaches, we have created a new ontology of strategies for treating IEMs and categorized treatments using this ontology. The use of an ontology allows for data integration, aggregation, and query expansion, as well as enhancing data access according to the FAIR (Findable, Accessible, Interoperable, Reusable) data principles [15]. We used a wide variety of public online data sources and, where possible, we extracted the specific aspects of phenotype addressed by the therapy from the literature through manual expert curation. DDIEM is freely available at http://ddiem.phenomebrowser.net.

## Results

DDIEM is a database of therapeutic approaches to treating IEMs which have been manually extracted from the scientific literature. DDIEM currently covers 300 rare diseases along with the association of 305 genes and 584 drugs that were used to treat these diseases; these treatment attempts have been linked to 1,482 distinct disease-associated phenotypes that were influenced by these treatments. The main entity in DDIEM is the therapeutic procedure which is used to treat a metabolic disorder; these procedures are classified based on their underlying mechanism or modality. We have developed an ontology using the Web Ontology Language (OWL) [16] to formalize this classification of treatments in DDIEM.

### Classification of treatment mechanisms: the dDIEM ontology

There are several existing strategies for classifying treatments for inborn errors of metabolism. Most of these are concerned with the underlying biochemical error and affected pathways. For example, ICD-11, based on the recommendations of the Society for the Study of Inborn Errors of Metabolism (SSIEM) [17], classifies inborn errors of metabolism mainly by type of metabolism; e.g., lipid metabolism, carbohydrate metabolism. ICD-11 also has other general axes of classification for IEMs, such as peroxisomal diseases, and more specific axes of classification, such as classifying diseases based on involvement in purine metabolism. Other nosologies emphasize diagnostic and clinical aspects of disease and, for example, classify diseases into disorders that either involve only one functional system (such as the immune system) or diseases in which the basic biochemical lesion affects metabolic pathways common to a large number of tissues [17]. The latter group is divided by molecule type of interest and energy metabolism. More recently, a nosology of IEM has been developed which classifies diseases by biochemical entity and process, a work that emerged from the European Reference Network for Hereditary Metabolic Disorders [8]. At least one classification is based neither on biochemical pathway nor clinical manifestation, but rather on modalities of treatment [12, 18].

While our major interest concerns medical treatment, we also include dietary modification (i.e., excluding or augmenting the diet with naturally occurring food substances). However, most of the available therapeutic options for IEMs are symptomatic where they treat the

symptoms or downstream effects without addressing the direct effects of the altered protein underlying the disease. Some other approaches include Enzyme Replacement Therapy (ERT) where the missing enzyme is replaced by infusions of an enzyme that is purified from human or animal tissue or blood or produced by novel recombinant techniques [19]. Typically, the enzyme is modified to allow for a longer half-life, more potent activity, resistance to degradation or targeting to a specific organ, tissue or cell type.

Some treatment strategies deliver their effects only if there is some residual enzymic activity and therefore their efficacy is dependent on this. Examples are substrate reduction therapy (SRT) [11], chemical chaperone therapy [20] and pharmacological chaperone therapy [21]. SRT is a strategy that works through limiting the amount of substrate synthesized to a level that can be effectively cleared by the impaired enzyme. The efficacy of SRT is mutation-specific and dependent on residual enzyme activity level [22]. Chaperone therapy uses small molecule substances, often osmolytes. These either facilitate folding or stabilize misfolded proteins to rescue residual activity, for example by reducing abnormal aggregation or interacting with active sites. [20]. In some cases a disease is amenable to direct gene therapy [23] which usually alters the somatic genome to prevent or treat a disease through insertion of a functional copy of the affected gene [20]. More recently, the potential for gene editing and epigenetic modifications are showing considerable promise [24].

We have developed and implemented the DDIEM ontology as a framework for classifying treatments by mechanism of action, as described above. This ontology allows the generation of a well-structured dataset broadly interoperable with existing resources. We based the DDIEM ontology on the Ontology of General Medical Sciences (OGMS) [25] and have made it available in the community ontology repositories BioPortal [26] and AberOWL [27].

In the DDIEM ontology, we classify mechanisms of treatment using three upper level classes. The first class covers those treatments that attempt to compensate for or modulate the biological functions affected by the dysfunctional protein (`mechanistically predicated therapeutic procedure`). The second class covers those that treat symptoms (`symptomatic therapeutic procedure`); and the third class covers surgical or physical procedures such as stem cell transplantation (`surgical or physical therapeutic procedure`). Some therapeutic strategies combine multiple drugs or other treatment modalities which work through different mechanisms, for example one drug addressing the primary lesion and another its symptomatic consequences. These combination therapies are also included as a separate class in the ontology. Figure 1 depicts the structure of the DDIEM

ontology; the definition of each class can be found in Table 1. Table 2 shows several examples of therapeutic procedure for the main classes in DDIEM.

The DDIEM ontology is based on an upper-level ontology for medical sciences, the OGMS. Reusing such an upper-level ontology allows us to consistently integrate the DDIEM ontology with related efforts of formalizing categories in the biomedical domain, including the Drug Ontology [28] and the Medical Action Ontology [29] which is currently under development and will broadly characterize therapeutic procedures.
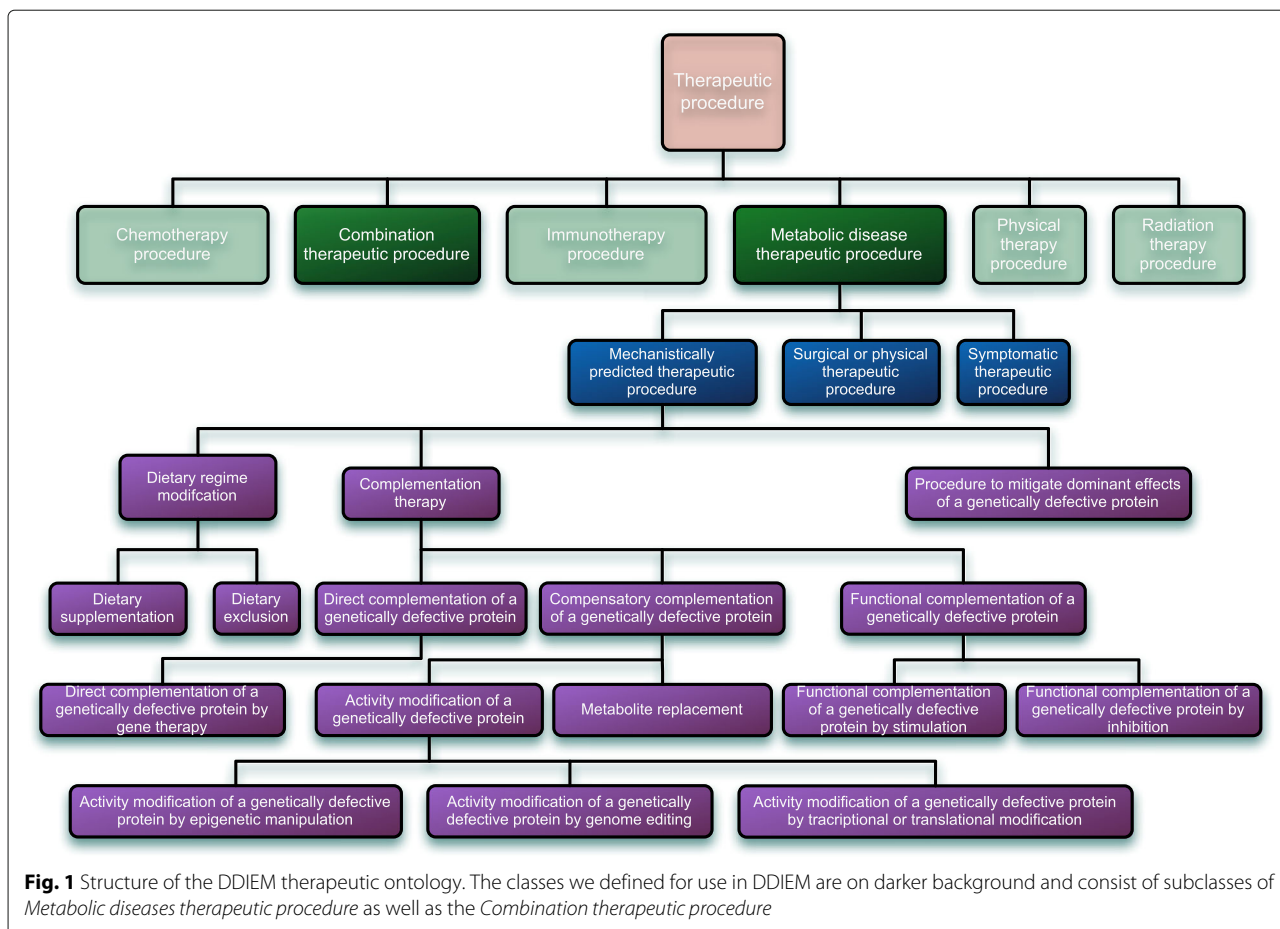
## DDIEM database

In DDIEM, we manually curated over 1,600 scientific literature articles to record evidence for therapies that have been attempted for currently 300 rare metabolic diseases; these 300 metabolic diseases are associated with 305 genes. We classify therapeutic approaches that have been reported in literature for each disease in DDIEM using the DDIEM ontology. The distribution of the drugs used as part of therapeutic procedures and their classification using the DDIEM ontology is shown in Fig. 2. These therapeutic approaches involve 584 unique drugs and correct or ameliorate 1,482 distinct phenotypes. For each therapeutic strategy, we record both evidence and provenance information using standardized identifiers and coding systems. Figure 3 provides an overview of the type of information we collect and the relations between the different types of information.

We identified 30 diseases where the outcomes of a therapeutic procedure were affected by particular genotypes or genetic variants. Modifying genotypes may include genotypes that affect the pharmacological action of a substance, either positively or negatively. For example, in familial hyperinsulinemic hypoglycemia type 3 (OMIM:602485), the Y214C variant in the *GCK* gene was found to result in patients' being unresponsive to diazoxide while patients carrying the M197I variant in the same gene did respond to the treatment [30]. When available, we record this information in DDIEM.

There are 55 orphan diseases in DDIEM for which we did not assign any therapeutic strategy class, either due to absence of information on tested therapeutic interventions in the literature, or the benign status of the disease which does not usually require any intervention. In the situation where we were unable to identify any report of a therapeutic strategy for a disease in DDIEM, we mark the disease as "no treatment is available".

Phenotypes in DDIEM that are affected by a therapeutic procedure are formally coded using the Human Phenotype Ontology (HPO) [31] or, if no HPO class could be identified, using the Mammalian Phenotype Ontology (MPO) [32]. In 113 cases we could not identify a phenotype class matching the described phenotype in

**Fig. 1** Structure of the DDIEM therapeutic ontology. The classes we defined for use in DDIEM are on darker background and consist of subclasses of *Metabolic diseases therapeutic procedure* as well as the *Combination therapeutic procedure*

these two ontologies and recorded the phenotype as free text; additionally, we requested extension of the HPO with the missing phenotypes so that we can formally include them in DDIEM once they become available in the HPO.

DDIEM includes reviewed reports for a therapeutic intervention; however, the quality of the evidence for a reported therapeutic intervention will differ between publications, ranging from individual case reports to large clinical trials. In DDIEM, we distinguish between six different types of evidence provided for the effect of a therapeutic procedure on disease-associated phenotypes, and we use the Evidence and Conclusion Ontology [47] to record evidence for our assertions. We distinguish evidence based on animal models (ECO:0000179), clinical trials ECO:0007121), and experiments on cell lines (ECO:0001565). In some cases, study authors suggested therapeutic procedures based on clinical observation alone; we record these as inference by a study author (ECO:0007764). In rare cases, DDIEM curators inferred information about a therapeutic procedure although it was not explicitly stated in the article; we mark this using the "inferred by curator" evidence (ECO:0000305).

Figure 4 shows the distribution of the evidence codes in DDIEM. While the evidence codes can provide information about the type of study that investigates a therapeutic intervention, they do not provide quantitative information such as sample sizes and statistical measures which are also required to evaluate the evidence for a therapeutic intervention; in the future, we may extend DDIEM to also include some quantitative information as part of our evidence model.

We will continue to extend DDIEM in the future both in its coverage of diseases as well as by updating information on diseases already included in DDIEM. For this purpose, we plan to search literature for diseases already included in DDIEM in frequent intervals, and the DDIEM curators will select and add new information as it becomes available.

## Implementation of fAIR principles

DDIEM is intended as a resource for biomedical and clinical researchers as well as for computational scientists. To enable DDIEM content to be usable by a wide range of researchers, we aim to follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable) [15].

**Table 1** A summary of the main mechanisms of therapeutic procedures for metabolic diseases

| Therapeutic procedure | Definition |
| --- | --- |
| Combination therapeutic procedure | A therapeutic procedure using multiple therapies to treat a single disease. Often all the therapies are pharmaceutical. It can also involve non-drug therapy, such as the combination of medications, behaviour modification, or physical procedures. The different components might act through the same or different mechanisms. |
| Mechanistically predicated therapeutic procedure | A therapeutic procedure which directly addresses the effects of a protein defect produced by the mutation, its structure, activity, or end product. Therapies which target the local interaction or regulatory network or pathway in which the affected protein lies are also regarded as mechanistic therapies. |
| Complementation therapy | A therapeutic procedure in which a treatment directly or indirectly compensates for the loss or gain of activity of a genetically defective protein within the network or pathway of which that protein is a member. |
| Direct complementation of a genetically defective protein | A therapeutic procedure in which a genetically defective protein is replaced by a canonical source of the same protein, genetically as in gene therapy, or by some other means of delivery, whose completion is hypothesized by a health care provider to eliminate a disorder or to alleviate the signs and symptoms of a disorder or pathological process. |
| Compensatory complementation of a genetically defective protein | A therapeutic procedure in which the availability, activity, stability, or turnover of a defective enzyme is modified by delivery of small or macro-molecules, or genetic or epigenetic manipulation. |
| Functional complementation of a genetically defective protein | A therapeutic procedure in which the composition or metabolic activity of the pathway or network in which the defective protein is found is modified, compensating for alteration of activity of that protein. |
| Dietary regime modification | A therapeutic procedure in which the diet is supplemented with or depleted in molecules closely related to the products or end processes of a genetically defective protein which occur as natural products in the diet. |
| Dietary exclusion | A therapeutic procedure in which the diet is depleted in molecules closely related to the products or end processes of a genetically defective protein, which occur naturally in the diet. |
| Dietary supplementation | A therapeutic procedure in which the diet is supplemented with foodstuffs containing molecules closely related to the products or end processes of a genetically defective protein, which occur naturally in the diet. |
| Metabolite replacement | A therapeutic procedure involving enteral, parenteral, or transdermal provision of small molecules closely related to the products or end processes of a genetically defective protein. |
| Activity modification of a genetically defective protein | A therapeutic procedure in which generally small molecules are delivered to the organism in order to directly increase, decrease, or alter the activity or stability of a genetically defective protein. |
| Symptomatic therapeutic procedure | A therapeutic procedure aimed at amelioration of one or more abnormal phenotypes generated as a consequence of a defective protein or process by a means unrelated to the immediate pathway or network environment of the defective protein. Generally working at the level of the tissue or overall organismal physiology. |
| Surgical or physical therapeutic procedure | A therapeutic procedure to mitigate the immediate or future effects of the presence of a genetically defective protein which involves physical conditioning or anatomical modification. |
| Functional complementation of a genetically defective protein by inhibition | A therapeutic procedure that inhibits a component of the pathway or network in which a genetically defective protein is found, compensating for alteration of activity of that protein. |
| Functional complementation of a genetically defective protein by stimulation | A therapeutic procedure that stimulates a component of the pathway or network in which a genetically defective protein is found, compensating for alteration of activity of that protein. |

**Table 2** Examples of the main mechanisms of therapeutic procedures for metabolic diseases

| Therapeutic procedure | Examples |
| --- | --- |
| Activity modification of a genetically defective protein | In Bartter syndrome type 4A (OMIM:602522), tanespimycin (17-allylamino-17-demethoxygeldanamycin or 17-AAG), an Hsp90 inhibitor, enhances the plasma membrane expression of mutant barttins (R8L and G47R) in Madin–Darby canine kidney cells [33]. |
| Activity modification of a genetically defective protein by epigenetic manipulation | HDACi 109/RG2833 increases FXN mRNA levels and frataxin protein, with concomitant changes in the epigenetic state of the gene, in the treatment of Friedreich's ataxia (OMIM:229300) [34]. |
| Activity modification of a genetically defective protein by genome editing | CRISPR-Cas9-mediated gene editing is a potential treatment for Hurler syndrome (OMIM:607014) as established in cell and animal based studies [35], [36]. |
| Activity modification of a genetically defective protein by transcriptional or translational modification | Vitamin D in treating hyperprolinemia (OMIM:239500) and recombinant human erythropoietin (rhuEPO) in treating Friedreich's ataxia (OMIM:229300) are examples of drugs that enhance the activity of PRODH and FXN respectively by transcriptional modulation of PRODH in hyperprolinemia and increasing frataxin expression in Friedrich's ataxia [37] [38]. |
| Direct complementation of a genetically defective protein | Recombinant human IGF1 (mecasermin) is a form of enzyme replacement therapy in treating insulin-like growth factor I deficiency (OMIM:608747) [39]. |
| Direct complementation of a genetically defective protein by gene therapy | An adeno-associated viral vector containing a porphobilinogen deaminase gene is a treatment for acute intermittent porphyria (OMIM:176000) [40]. |
| Functional complementation of a genetically defective protein by inhibition | Miglustat is a form of substrate reduction therapy (SRT) used to treat Gaucher's disease (OMIM:230800) where miglustat inhibits the ceramide-specific glucosyltransferase which catalyses the first committed step of GSL synthesis [41]. |
| Procedure to mitigate dominant effect of a genetically abnormal protein | In treating hyperinsulinism-hyperammonemia syndrome (OMIM:606762), diazoxide inhibits insulin release and promotes complete resolution of hypoglycemia [42]. |
| Dietary exclusion | A valine-restricted diet is a treatment for patients suffering from isobutyryl-CoA dehydrogenase deficiency (OMIM:611283), which inolves valine metabolism [43]. |
| Dietary supplementation | Beneprotein and uncooked cornstarch as high sources of protein and carbohydrates are used to prevent long-term complications in glycogen storage disease IXb (OMIM:261750) patients [44]. |
| Metabolite replacement | Hydrocortisone is a replacement therapy for patients with 17-alpha-hydroxylase deficiency (OMIM:202110) [45]. |
| Functional complementation of a genetically defective protein by stimulation | Sodium valproate activates the expression of one glycogen phosphorylase isoform, GP-BB, which in turn results in a decrease in intracellular glycogen accumulation, a dominant feature of glycogen storage disease type V (OMIM:232600) [46]. |

Specifically, to ensure interoperability, we followed the Linked Data principles [48] and linked the entities described in DDIEM to community reference resources and ontologies. Diseases are mapped to Online Mendelian Inheritance in Man (OMIM) [49] disease identifiers as well as to identifiers in the IEMbase database [10]. Genes are linked to the NCBI Entrez gene database [50] and their products to Expasy [51], KEGG [52], and the UniProt [53] databases. We mapped drugs to Drugbank [54], PubChem [55], ChEBI [56], and to identifiers from the Anatomical Therapeutic Chemical Classification System (ATC). In DDIEM, we rely on ontologies in the OBO Foundry [57] as collaboratively developed reference ontologies in the biomedical domain. We represent phenotypes using either the Human Phenotype Ontology (HPO) [31] or the Mammalian Phenotype Ontology (MPO) [32], and we use the Evidence and Conclusion Ontology (ECO) [47] to specify different study types and evidences.

DDIEM content is accessible through a website as well as through a public SPARQL end-point to enable computational access. DDIEM data is also downloadable, and each release of the data is assigned a unique Digital Object Identifier (DOI) [59]. To make DDIEM content findable, we registered DDIEM on the FAIRsharing platform [60] and the DDIEM ontology in several ontology repositories.
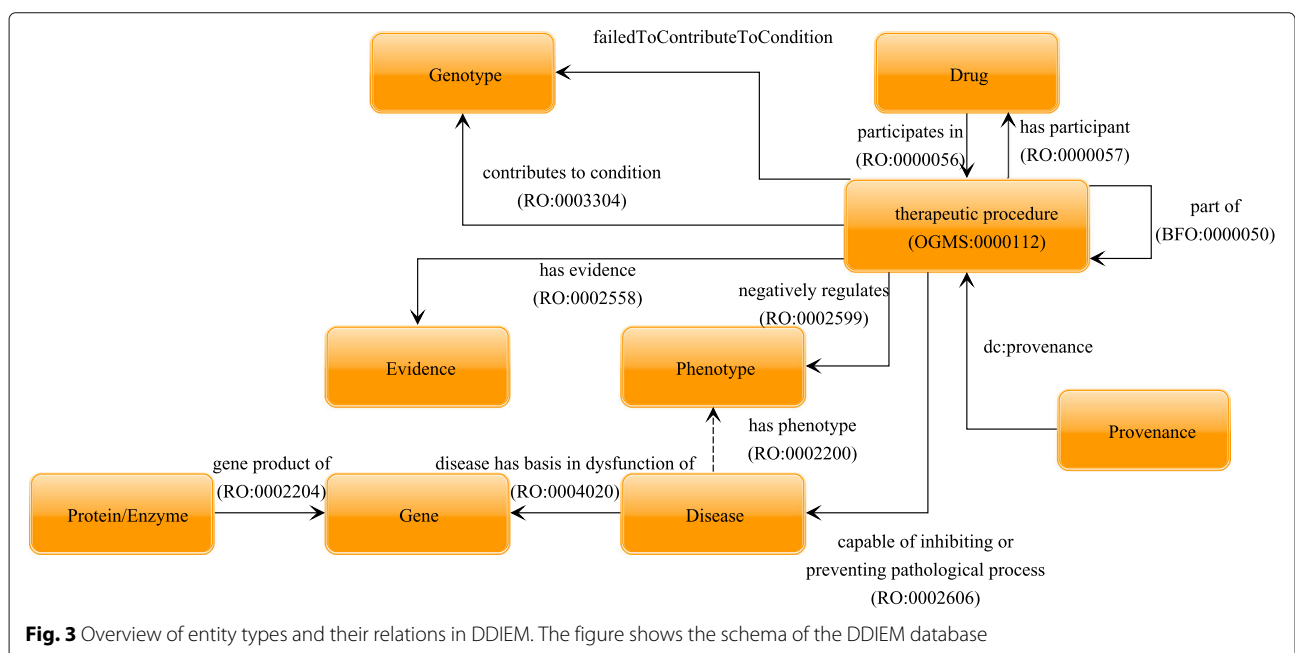
## Discussion

There exist several formal data resources for information on therapies for rare metabolic diseases. However, the majority of them focus on symptoms, clinical or metabolic

**Fig. 2** Distribution of the therapeutic procedures in DDIEM based on their classes in the DDIEM ontology
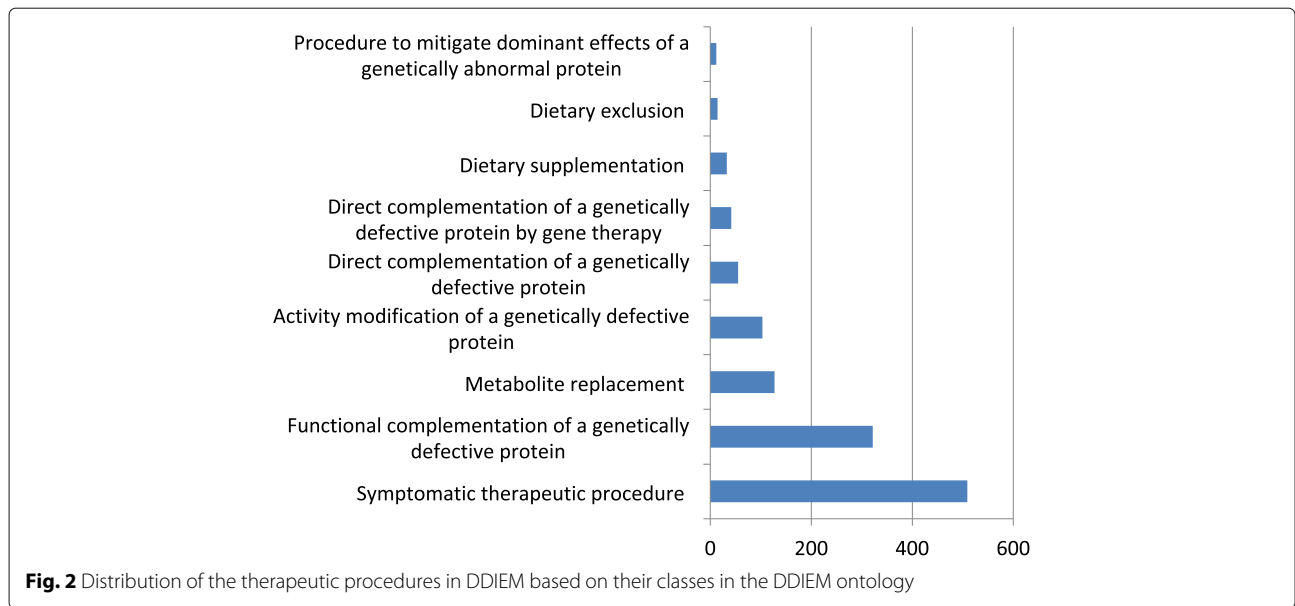
aspects rather than therapeutic approaches. For example, the Orphanet [14] database includes drugs designated for treating rare genetic diseases. Similarly, RAMEDIS [9] focuses on treatments described in case reports for a number of metabolic diseases. However, neither provide information on the mechanism of drug action or on the phenotypes corrected or alleviated. Another database, IEMbase [10], provides some therapeutic information but concentrates mainly on the clinical and biochemical aspects of rare metabolic disease, and provides an expert platform to facilitate their early and accurate diagnosis. To the

best of our knowledge, DDIEM is the first database that significantly focuses on the treatment to correct or alleviate phenotypes associated with the course of a specific metabolic disease.

However, the main difference between DDIEM and other databases is its focus on collecting all reports that provide any evidence for the use of a therapeutic intervention on an IEM, independently of whether the therapy was successful or not. DDIEM therefore provides information primarily for research use, in particular for research involving drug repurposing approaches where
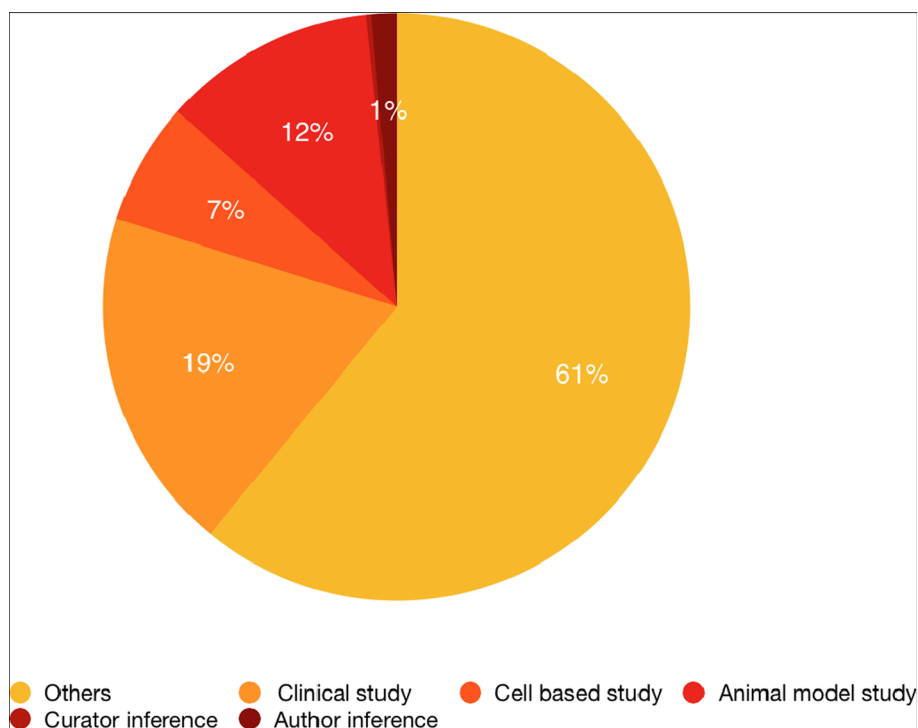


**Fig. 3** Overview of entity types and their relations in DDIEM. The figure shows the schema of the DDIEM database

**Fig. 4** Distribution of evidence codes used to annotate the supporting evidence for therapeutic procedures included in DDIEM

the negative cases may also provide insights into molecular mechanisms, in particular when used as part of machine learning approaches. When using DDIEM for the purpose of drug repurposing, it is important to consider that the negative cases will likely suffer from reporting bias [61].

With a focus on phenotypes and mechanisms, DDIEM uniquely addresses fundamental aspects of rare disease therapy with the aim of supporting the analysis of the therapeutic landscape and the development of new treatments. The Data Mining and Repurposing (DMR) task force of the IRDiRC [62] recently emphasized the important contribution of data mining and data integration to the development of new drugs for rare diseases, and highlighted the need to get data out of silos – often semantic silos – to facilitate in silico approaches to drug development and repurposing. DDIEM is developed on the FAIR principles to ensure findability, accessibility, interoperability, and reusability and is accessible either through a web interface or computationally through the DDIEM API. Use of community standards, databases, and ontologies permits ready computational integration of curated DDIEM information into other datasets. DDIEM therefore provides a dataset that can be reused by a wide range of researchers, using different methodological approaches, to investigate existing and develop new therapeutic approaches.

The current data in DDIEM does not cover the details or scale of clinical trials or individual reports, the drug formulation, or the dosage applied. Such information could further provide valuable information particularly for drug repurposing. In the future, we plan to expand DDIEM by adding a limited amount of quantitative information and dosage, as well as reporting quantitative drug effects.

## Conclusion

We developed DDIEM, a database focusing on treatments for Inborn Errors of Metabolism. DDIEM integrates literature-reported treatments for a wide range of IEMs together with information on the outcome. The reports range from individual case reports to clinical trials. Consequently, DDIEM provides information on the on- and off-label uses of drugs and their effects for treating IEMs and may be used to accelerate drug development and drug repurposing for these diseases.

In creating DDIEM we have extensively analyzed publicly available data on the treatment of rare metabolic diseases, and present it in a way which permits integration with other datasets and resources. The value we add to the currently available public data is in our expert curation and semantic formalization, integrating and presenting this data in a way compliant with FAIR standards, and making it freely accessible to investigators in both public and private domains. DDIEM is accessible

through its website and can be queried computationally. We support the conclusions and recommendations of the IRDiRC regarding data integration and exploitation [62], and believe that DDIEM will contribute to the discovery or repurposing of drugs for diseases still lacking effective therapies.

## Materials and methods
### Resources used
We rely on literature resources listed in PubMed and information published on ClinicalTrials.gov to gather information about treatments that were used for rare metabolic diseases. For parts of the curation process we used the GoPubMed software [63] to access and query literature resources. While PubMed reports on rather complete studies, ClinicalTrials.gov provides information for ongoing studies with potential interim results for each rare disease. We obtained a list of metabolic diseases from the Genetic and Rare Diseases (GARD) information center database [64].

### Literature curation
To create the DDIEM content, we searched in literature the disease names or synonyms from the metabolic disease list we gathered from GARD. In the cases where we could not find any literature records for the name or synonyms of a given disease, we used the gene name associated with the disease in the search. We then filtered all the retrieved records by selecting the ones containing treatment or management information. Since these diseases are considered rare, we did not exclude any literature as long as it provides information about therapeutic interventions, and in which we found explicit evidence for the correlation between the applied treatment and clinical phenotypes, which are either completely treated or alleviated. However, in DDIEM, we focused mainly on pharmacological and, to some extent, on dietary interventions. We then curated the resulting literature documents manually to extract the metabolic disorder investigated, the drugs used to treat the disorder, the effects of the treatment, and the suggested mechanism of action. If drugs exerted their impact differently between patients with the same metabolic disease due to different genetic variants in the same gene, we added those variants as an additional reference. We categorized variants that affect the effectiveness and success of therapeutic intervention into two categories based on whether they promoted the drug effect or not.

We provided all literature references as provenance information through the DDIEM web interface. After mapping most drugs to their identifiers in the cross-linked databases, there were some therapies we could not link to any database. Hence, we indicated their identifiers as non-available (NA).

Additionally, we developed a therapeutic ontology that describes the mechanism by which the therapeutic procedures act to either adequately treat the disease or modify disease pathogenesis. We recorded the disease phenotypes which were treated or improved by the therapeutic procedure and mapped them to reference ontologies. Similarly, where phenotypes had no matching identifiers in those ontologies, we referred to identifiers as non-available (NA).

### Data representation and web interface
We represent the content of DDIEM using the Resource Description Framework (RDF) [65]. First, we normalized the curated data (which is primarily stored in a comma-separated values text file format) by mapping the DDIEM content to their identifiers in different databases (Uniprot, Expasy, KEGG, OMIM and Drugbank). In the second step, we represent the DDIEM content using the DDIEM RDF data model which captures the relations between the biomedical entitites we characterize in DDIEM.

We developed a web interface to allow users to navigate the DDIEM database. The interface provides a list of diseases covered in the database along with their details including treatments, participating drugs, phenotypes, and references to the publications from which the information was extracted. The web interface also provides a search in the DDIEM resources using disease, drug, and mode of action of the therapeutic procedure. Furthermore, to enable automated access to the DDIEM content, we provide a SPARQL endpoint which directly queries the RDF data we provide.

We further implemented a web server using Node.js and developed an API for accessing DDIEM resources. The API response is formatted in either JSON or JSON-LD format which enables the possibility to build client applications using the RDF data underlying DDIEM. We used OpenLink's Virtuoso database as an RDF store to store and query the DDIEM data, and all primary data is stored in the RDF store and the Node.js server retrieves all data through SPARQL queries. The DDIEM data is also available for download in RDF format from the DDIEM website at http://ddiem.phenomebrowser.net.

## Abbreviations
ChEBI: Chemical Entities of Biological Interest; DDIEM: Drug database of inborn errors of metabolism; DMR: Data Mining and Repurposing; ECO: Evidence and Conclusion Ontology; ERT: Enzyme replacement therapy; FAIR: Findable, Accessible, Interoperable, Reusable; FDA: Food and Drug Administration; GARD: Genetic and Rare Diseases Information Center; HPO: Human Phenotype Ontology; ICD: International Classification of Diseases; IEM: Inborn errors of metabolism; IRDiRC: International Rare Diseases Research Consortium; KEGG: Kyoto Encyclopedia of Genes and Genomes; MAxO: Medical Action Ontology; MPO: Mammalian Phenotype Ontology; NA: Not available; NCBI: National Center for Biotechnology Information; NIH: National Institutes of Health; OGMS: Ontology for General Medical Science; OMIM: Online Mendelian Inheritance in Man; OWL: Web Ontology Language; RAMEDIS: Rare Metabolic Diseases Database; RDF: Resource Description Framework; SRT: Substrate

## Availability of data and materials
DDIEM is freely available at http://ddiem.phenomebrowser.net. Source code and all data used to construct DDIEM is also freely available on https://github.com/bio-ontology-research-group/DDIEM.

## Ethics approval and consent to participate
Not applicable.

## Consent for publication
Not applicable.

## Competing interests
The authors declare that they have no competing interests.

## Author details
[1]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, 4700 KAUST, 23955 Thuwal, Kingdom of Saudi Arabia. [2]Computer, Electrical and Mathematical Sciences & Engineering Division (CEMSE), King Abdullah University of Science and Technology, 4700 KAUST,PO 23955 Thuwal, Saudi Arabia. [3]Department of Physiology, Development & Neuroscience, University of Cambridge, Downing Street, CB2 3EG Cambridge, United Kingdom.

## References
1. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. Lancet. 2008;371(9629):2039–41.
2. Austin CP, Cutillo CM, Lau LPL, Jonker AH, Rath A, Julkowska D, Thomson D, Terry SF, de Montleau B, Ardigo D, Hivert V, Boycott KM, Baynam G, Kaufmann P, Taruscio D, Lochmuller H, Suematsu M, Incerti C, Draghia-Akli R, Norstedt I, Wang L, Dawkins HJS, International Rare Diseases Research Consortium (IRDiRC). Future of rare diseases research 2017-2027: An irdirc perspective. Clin Transl Sci. 2018;11(1):21–7. https://doi.org/10.1111/cts.12500.
3. Sun W, Zheng W, Simeonov A. Drug discovery and development for rare genetic disorders. Am J Med Genet A. 2017;173(9):2307–22. https://doi.org/10.1002/ajmg.a.38326.
4. Braun MM, Farag-El-Massah S, Xu K, Cote TR. Emergence of orphan drugs in the united states: a quantitative assessment of the first 25 years. Nat Rev Drug Discov. 2010;9(7):519–22. https://doi.org/10.1038/nrd3160.
5. The Committee for Orphan Medicinal Products and the European Medicines Agency Scientific Secretariat, Westermark K, Holm BB, Soderholm M, Llinares-Garcia J, Riviere F, Aarum S, Butlen-Ducuing F, Tsigkos S, Wilk-Kachlicka A, N'Diamoi C, Borvendeg J, Lyons D, Sepodes B, Bloechl-Daum B, Lhoir A, Todorova M, Kkolos I, Kubackova K, Bosch-Traberg H, Tillmann V, Saano V, Heron E, Elbers R, Siouti M, Eggenhofer J, Salmon P, Clementi M, Krievins D, Matuleviciene A, Metz H, Vincenti AC, Voordouw A, Dembowska-Baginska B, Nunes AC, Saleh FM, Foltanova T, Mozina M, Torrent i Farnell J, Beerman B, Mariz S, Evers MP, Greene L, Thorsteinsson S, Gramstad L, Mavris M, Bignami F, Lorence A, Belorgey C. European regulation on orphan medicinal products: 10 years of experience and future perspectives. Nat Rev Drug Discov. 2011;10(5):341–9. https://doi.org/10.1038/nrd3445.
6. Miller KL, Lanthier M. Investigating the landscape of us orphan product approvals. Orphanet J Rare Dis. 2018;13(1):183. https://doi.org/10.1186/s13023-018-0930-3.
7. Elliott E, Zurynski Y. Rare diseases are a 'common' problem for clinicians. Aust Fam Physician. 2015;44:630–3.
8. Ferreira CR, van Karnebeek CDM, Vockley J, Blau N. A proposed nosology of inborn errors of metabolism. Genet Med. 2019;21(1):102–6. https://doi.org/10.1038/s41436-018-0022-8.
9. Töpel T, Hofestädt R, Scheible D, Trefz F. Ramedis. Appl Bioinformatics. 2006;5(2):115–8. https://doi.org/10.2165/00822942-200605020-00006.
10. Lee JJY, Wasserman WW, Hoffmann GF, van Karnebeek CDM, Blau N. Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. Genet Med. 2018;20(1):151–8. https://doi.org/10.1038/gim.2017.108.
11. Yue WW, Mackinnon S, Bezerra GA. Substrate reduction therapy for inborn errors of metabolism. Emerg Top Life Sci. 2019;20180058:. https://doi.org/10.1042/ETLS20180058.
12. Gambello MJ, Li H. Current strategies for the treatment of inborn errors of metabolism. J Genet Genomics. 2018;45(2):61–70. https://doi.org/10.1016/j.jgg.2018.02.001.
13. Ginocchio VM, Brunetti-Pierri N. Progress toward improved therapies for inborn errors of metabolism. Hum Mol Genet. 2016;25(R1):27–35. https://doi.org/10.1093/hmg/ddv418.
14. Weinreich SS, Mangon R, Sikkens J, Teeuw M, Cornel M. Orphanet: a european database for rare diseases. Ned Tijdschr Geneeskd. 2008;152(9):518–9.
15. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The fair guiding principles for scientific data management and stewardship. Scientific Data. 2016;3(1):160018. https://doi.org/10.1038/sdata.2016.18.
16. Grau BC, Horrocks I, Motik B, Parsia B, Patel-Schneider P, Sattler U. Owl 2: The next step for owl. Web Semantics: Science, Services and Agents on the World Wide Web. 2008;6(4):309–322. https://doi.org/10.1016/j.websem.2008.05.001.
17. Zschocke J. In: Blau N, Duran M, Gibson KM, Dionisi Vici C, editors. SSIEM Classification of Inborn Errors of Metabolism. Berlin, Heidelberg: Springer; 2014, pp. 817–830.
18. Das AM, Illsinger S. In: Hock FJ, Gralinski MR, editors. Pharmacological Therapy in Inborn Errors of Metabolism. Cham: Springer; 2018, pp. 1–16.
19. Serrano J. Livertox: An online information resource and a site for case report submission on drug-induced liver injury. Clin Liver Dis. 2014;4(1):22–5.
20. Haneef SAS, Doss CGP, Vol. 102. In: Donev R, editor. Advances in Protein Chemistry and Structural Biology. Cambridge: Academic Press; 2016, pp. 225–265. Chap. 8 - Personalized Pharmacoperones for Lysosomal Storage Disorder.
21. Small SA. Pharmacological chaperones in the age of proteomic pathology. Proc Natl Acad Sci U S A. 2014;111(34):12274–5. https://doi.org/10.1073/pnas.1413111111.
22. El-Gharbawy A, Vockley J. In: Jefferies JL, Blaxall BC, Robbins J, Towbin JA, editors. Cardioskeletal Myopathies in Children and Young Adults. Cambridge: Academic Press; 2017, pp. 265–303. Chap. 14 - Nonmitochondrial Metabolic Cardioskeletal Myopathies.
23. Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K, Sadelain M. Gene therapy comes of age. Science. 2018;359(6372):. https://doi.org/10.1126/science.aan4672.

24. Rutten MGS, Rots MG, Oosterveer MH. Exploiting epigenetics for the treatment of inborn errors of metabolism. J Inherit Metab Dis. 2019. https://doi.org/10.1002/jimd.12093.

25. Scheuermann R, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit Transl Bioinform. 2009;2009:116–20.

26. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey MA, Chute CG, Musen MA. Bioportal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res. 2009;37(Web Server issue):170–3. https://doi.org/10.1093/nar/gkp440.

27. Hoehndorf R, Slater L, Schofield PN, Gkoutos GV. Aber-owl: a framework for ontology-based data access in biology. BMC Bioinformatics. 2015;16:26.

28. Hogan WR, Hanna J, Joseph E, Brochhausen M. Towards a consistent and scientifically accurate drug ontology. In: CEUR Work Proc; 2013. p. 68–73.

29. MAxO. https://github.com/monarch-initiative/MAxO. Accessed 6 Jan 2020.

30. Sayed S, Langdon DR, Odili S, Chen P, Buettger C, Schiffman AB, Suchi M, Taub R, Grimsby J, Matschinsky FM, et al. Extremes of clinical and enzymatic phenotypes in children with hyperinsulinism caused by glucokinase activating mutations. Diabetes. 2009;58(6):1419–27.

31. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, Brudno M, Buske OJ, Chinnery PF, Cipriani V, Connell LE, Dawkins HJS, DeMare LE, Devereau AD, de Vries BA, Firth HV, Freson K, Greene D, Hamosh A, Helbig I, Hum C, Jähn JA, James R, Krause R, F Laulederkind SJ, Lochmüller H, Lyon GJ, Ogishima S, Olry A, Ouwehand WH, Pontikos N, Rath A, Schaefer F, Scott RH, Segal M, Sergouniotis PI, Sever R, Smith CL, Straub V, Thompson R, Turner C, Turro E, Veltman MWM, Vulliamy T, Yu J, von Ziegenweidt J, Zankl A, Züchner S, Zemojtel T, Jacobsen JOB, Groza T, Smedley D, Mungall CJ, Haendel M, Robinson PN. The human phenotype ontology in 2017. Nucleic Acids Res. 2017;45(D1):865–76. https://doi.org/10.1093/nar/gkw1039.

32. Smith CL, Eppig JT. Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. J Biomed Semantics. 2015;6:11. https://doi.org/10.1186/s13326-015-0009-1.

33. Nomura N, Kamiya K, Ikeda K, Yui N, Chiga M, Sohara E, Rai T, Sakaki S, Uchida S. Treatment with 17-allylamino-17-demethoxygeldanamycin ameliorated symptoms of Bartter syndrome type iv caused by mutated bsnd in mice. Biochem Biophys Res Commun. 2013;441(3):544–9.

34. Soragni E, Miao W, Iudicello M, Jacoby D, De Mercanti S, Clerico M, Longo F, Piga A, Ku S, Campau E, et al. Epigenetic therapy for Friedreich ataxia. Ann Neurol. 2014;76(4):489–508.

35. de Carvalho TG, Schuh R, Pasqualim G, Pellenz FM, Filippi-Chiela EC, Giugliani R, Baldo G, Matte U. Crispr-cas9-mediated gene editing in human mps i fibroblasts. Gene. 2018;678:33–7.

36. Schuh RS, Poletto É, Pasqualim G, Tavares AMV, Meyer FS, Gonzalez EA, Giugliani R, Matte U, Teixeira HF, Baldo G. In vivo genome editing of mucopolysaccharidosis i mice using the CRISPR/Cas9 system. J Control Release. 2018;288:23–33.

37. Clelland JD, Read LL, Drouet V, Kelly A, Duff KE, Nadrich RH, Rajparia A, Clelland CL. Vitamin d insufficiency and schizophrenia risk: evaluation of hyperprolinemia as a mediator of association. Schizophr Res. 2014;156(1):15–22.

38. Sturm B, Stupphann D, Kaun C, Boesch S, Schranzhofer M, Wojta J, Goldenberg H, Scheiber-Mojdehkar B. Recombinant human erythropoietin: effects on frataxin expression in vitro. Eur J Clin Investig. 2005;35(11):711–7.

39. Camacho-Hübner C, Woods KA, Miraki-Moud F, Hindmarsh PC, Clark AJ, Hansson Y, Johnston A, Baxter RC, Savage MO. Effects of recombinant human insulin-like growth factor i (igf-i) therapy on the growth hormone-igf system of a patient with a partial igf-i gene deletion. J Clin Endocrinol Metab. 1999;84(5):1611–6.

40. Yasuda M, Bishop DF, Fowkes M, Cheng SH, Gan L, Desnick RJ. Aav8-mediated gene therapy prevents induced biochemical attacks of acute intermittent porphyria and improves neuromotor function. Mol Ther. 2010;18(1):17–22.

41. Lachmann RH, Platt FM. Substrate reduction therapy for glycosphingolipid storage disorders. Expert Opin Investig Drugs. 2001;10(3):455–66.

42. Corrêa-Giannella ML, Freire DS, Cavaleiro AM, Fortes MAZ, Giorgi RR, Pereira MAA. Hyperinsulinism/hyperammonemia (hi/ha) syndrome due to a mutation in the glutamate dehydrogenase gene. Arq Bras Endocrinol Metabologia. 2012;56(8):485–9.

43. Pedersen CB, Bischoff C, Christensen E, Simonsen H, Lund AM, Young SP, Koeberl DD, Millington DS, Roe CR, Roe DS, et al. Variations in ibd (acad8) in children with elevated c 4-carnitine detected by tandem mass spectrometry newborn screening. Pediatr Res. 2006;60(3):315.

44. Roscher A, Patel J, Hewson S, Nagy L, Feigenbaum A, Kronick J, Raiman J, Schulze A, Siriwardena K, Mercimek-Mahmutoglu S. The natural history of glycogen storage disease types vi and ix: long-term outcome from the largest metabolic center in canada. Mol Genet Metab. 2014;113(3):171–6.

45. Kim SM, Rhee JH. A case of 17 alpha-hydroxylase deficiency. Clin Exp Reprod Med. 2015;42(2):72–6.

46. de Luna N, Brull A, Guiu JM, Lucia A, Martin MA, Arenas J, Martí R, Andreu AL, Pinós T. Sodium valproate increases the brain isoform of glycogen phosphorylase: looking for a compensation mechanism in mcardle disease using a mouse primary skeletal-muscle culture in vitro. Dis Models Mech. 2015;8(5):467–72.

47. Giglio M, Tauber R, Nadendla S, Munro J, Dessimoz C, Olley D, Ball S, Mitraka E, Schriml LM, Gaudet P, Hobbs ET, Erill I, Siegele DA, Hu JC, Mungall C, Chibucos MC. ECO, the Evidence & Conclusion Ontology: community standard for evidence information. Nucleic Acids Res. 2019;47(Database issue):D1186–D1194. https://doi.org/10.1093/nar/gky1036.

48. Bizer C, Heath T, Berners-Lee T. Linked data – the story so far. Int J Semantic Web Inf Syst. 2009;5(3):1–22.

49. Amberger JS, Bocchini CA, Scott AF, Hamosh A. Omim.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res. 2019;47(D1):1038–43. https://doi.org/10.1093/nar/gky1151.

50. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez gene: gene-centered information at ncbi. Nucleic Acids Res. 2010;39(suppl1):52–7. https://doi.org/10.1093/nar/gkq1237.

51. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. Expasy: Sib bioinformatics resource portal. Nucleic Acids Res. 2012;40(W1):597–603. https://doi.org/10.1093/nar/gks400.

52. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. Kegg: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2016;45(D1):353–61. https://doi.org/10.1093/nar/gkw1092.

53. The UniProt Consortium. Uniprot: a worldwide hub of protein knowledge. Nucleic Acids Res. 2018;47(D1):506–15. https://doi.org/10.1093/nar/gky1049.

54. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. 2017;46(D1):1074–82.

55. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, et al. Pubchem substance and compound databases. Nucleic Acids Res. 2015;44(D1):1202–13.

56. Degtyarenko K, De Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M. Chebi: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2007;36(suppl_1):344–50.

57. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Leontis N, Serra PR, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotech. 2007;25(11):1251–5.

58. Chibucos MC, Siegele DA, Hu JC, Giglio M. The evidence and conclusion ontology (eco): Supporting go annotations. In: Dessimoz c, kunca n, editors. theš gene ontology handbook. Methods in molecular biology, 1446; 2016.

59. Abdelhakim M, McMurray E, Syed AR, Kafkas S, Kamau A, Schofield DP, Hoehndorf R. DDIEM - Drug Database for Inborn Errors of Metabolism. https://doi.org/10.5281/zenodo.3597933.

60. Sansone S.-A., McQuilton P., Rocca-Serra P., Gonzalez-Beltran A., Izzo M., Lister A. L., Thurston M., FAIRsharing Community. Fairsharing as a community approach to standards, repositories and policies. Nat Biotechnol. 2019;37:358–67.

61. Ioannidis JPA. Why most published research findings are false. PLOS Med. 2005;2(8):. https://doi.org/10.1371/journal.pmed.0020124.

62. Southall NT, Natarajan M, Lau LPL, Jonker AH, Deprez B, Guilliams T, Hunter L, Rademaker CM, Hivert V, Ardigò D. The use or generation of biomedical data and existing medicines to discover and establish new

treatments for patients with rare diseases–recommendations of the irdirc data mining and repurposing task force. Orphanet J Rare Dis. 2019;14(1):1–8.

63. Doms A, Schroeder M. Gopubmed: exploring pubmed with the gene ontology. Nucleic Acids Res. 2005;33(S2):783–6. https://doi.org/10.1093/nar/gki470.

64. National Center for Advancing Translational Sciences (NCATS). Genetic and Rare Diseases Information Center (GARD). 2011. https://rarediseases.info.nih.gov/. Accessed 09 Dec 2019.

65. Gandon F, Schreiber G. RDF 1.1 XML Syntax. W3C recommendation. Cambridge: W3C; 2004. https://www.w3.org/TR/rdf-syntax-grammar/.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.